Audiovisual Accessibility: Evaluating Workflows for Closed Captioning and Transcripts White Paper

Stefano Kinkade, Jeremy Myntti, and Molly Rose Steed

May 31, 2019

Researchers worldwide rely on the ability to access primary source material to conduct their research activities. While many forms of digital content are accessible to most people through the metadata describing the items or optical character recognition (OCR) text which has been automatically derived from the text on an image, audiovisual content is not as easily accessible if there is no accompanying transcript or captioning. If a researcher has a disability such as a hearing impairment, the audiovisual content that we have available is useless without an alternate way of access.

With a large amount of audiovisual content being made publically available, we need to ensure that we have the means for providing the best user experience for all of our users. According to University of Utah policy 4-003 VI.B.3, "The University of Utah is committed to the principle of universal access to official information and encourages Web page design for accessibility to audiences using alternate communication channels" (https://regulations.utah.edu/it/4-003.php). In order to conform to this policy, we have researched different methods for making our digital audiovisual content more accessible by providing transcripts and captioning services. Providing transcripts for audiovisual content not only allows us to make our digital collections more accessible for users with a disability, it also allows for increased access for other researchers who can use the data in digital humanities projects for text mining or visualizations.

Following are notes on the different workflows, services, and tools that were used for this project.

Manual Transcription Notes
(In order of chronological transcription)

Title	Notes
13_Employment_SpeedsUp	There are quite a few segments in this clip that are mumbled or generally unintelligible. Audio speeds up around the 10 minute mark, causing it to become more difficult to transcribe as it gets faster and faster. Right at about the 44:30 mark, the audio becomes so sped up its almost indecipherable. There is much more guesswork involved after that point when compared with the rest of the clip.
A0194_09.wav	Really good quality. Nice and easy to transcribe.

74-7-24-47_HeartThrobsOfTheWest_TheDorionWoman-WAV	The voices are quite soft when compared to the static that is much more predominant in this tape, makes it pretty difficult to transcribe.
A0095_HeverJSessions-WAV	No significant notes.
A0095_MaymeWellsLower-WAV	No significant notes.
A0409_0001_MaudMayBabcock_Final_Address	One of the easiest to transcribe thanks to the enunciation of the speaker. Noisey transfer hardly had any negative effects on the process.
A0055_Shakespeare_Hamlet_1962	Take note I am cross checking text with a .pdf of the Hamlet play I found online, as per the recommendation in the evaluation files. "Cross Checking" is also a very generous way of putting it, considering how muddy the audio is. The speech is extremely difficult to parse.
A0212_02_Seeing_is_Believing	No significant notes.
Gould 1992	No significant notes.
A0938_OLTapp_21_ThoseOtherYears	No significant notes.
A0832_Cambodia2001	No significant notes.
D-4_Steve_Lacy_Mini_DV	No significant notes.
37_Media_part1_PoorQuality	No significant notes.

Workflow summary

Overall, this workflow is the most straightforward of the three. It is also the least volatile, as there is no dependence on external sources to get the job done. Everything can be done in-house. This is also a drawback, of course, as the transcription is only as dependable as the number of people working on it. The speed at which transcripts are produced is largely dependant on the typing speed of the individual and the quality of the transcripts is also affected by the transcriptionist's grammatical prowess. Of the three methods, this is the most prone to human-error. Hiring a professional in-house transcriptionist would generate the best transcripts for this workflow.

This workflow is best suited towards low-quality clips, those that Rev.com refuses to work on and Google generates garbage for.

Google Transcript Correction Notes
(In order of chronological transcription)

Title	Notes
13_Employment_SpeedsUp._mono.flac.txt	Google seems to pick and choose who it wants to

	be good at transcribing. The clip features multiple speakers, the first required a lot of correction, but the second needed a lot less. Approximately 13 and a half minutes into the clip, the google transcription becomes nearly useless. It is actually more of a hassle to parce what section google transcribed than it would be to just type it out.
27_Homemakers_part1	At the 57:29 mark, the audio cuts to a completely different clip. The quality of the audio is much worse and was more difficult to transcribe.
A0212_02_Seeing_is_Believing	Corrections for google transcripts mostly involve grammar corrections when the clip is as clean and clear as this one. Every caption begins capitalized and google never adds punctuation unless the person specifically says "Period," or "Question mark," or similar words.
A0095_HeberJSessions-MP3	No significant notes.
74-7-24-47_HeartThrobsOfTheWest_TheDorionWoman-WAV	No significant notes.
1994_Oregon	Hey y'all, just me speaking from the heart here, MovieCaptioner may not be the best option for correcting transcription. Right now, it really only serves to preview captions and convert text files to .vtt and vice versa. If we manage to get another third party software to convert .vtt to plain text and back, we could ditch MovieCaptioner completely because VLC media player can totally read .vtt files, so that would solve both the preview and conversion problems.
A0668_DaveTatsuno	Files like this seem to be an almost perfect match for google. The speech is slow and steady, so there are very few errors.
A0095_MaymeWellsLower-MP3	No significant notes.
A0194_09	No significant notes.
A0303KUTV_1995_June_09_5PM_6PM_10PM	After scrubbing through it, I found about the first hour of this clip is almost impossible to parse, but after that things become easier to discern, so my actual transcription starts at the 1:01:32 mark.
Luna_Mesa_DVD (Incomplete)	Stopped transcription due to technical issues. It looks like there's something wrong intrinsically with the raw file.

A0055_Shakespeare_Hamlet_1962 (Incomplete)	The google generated transcripts for this one are almost completely worthless. The captions are sparratically placed, and even then they almost never correctly guess what is being said. It would be a better option to just do something like this manually and not waste the time or resources on generating a transcript. The audio is way too muddy for the current transcription technology to handle.
A0832_Cambodia2001	No significant notes.
A0938_OLTapp_21_ThoseOtherYears	Google Transcriptions are a perfect fit for this tape. Only minimal corrections are needed.
Gould 1992	Technically, because all the transcriptions are pre-written, this should go faster than the manual transcription, but the process is actually slower when compared to the manual transcription. The fact I need to take my hands off the keyboard to mouse down and select each caption and wait half a second is definitely the bottleneck.
D-4_Steve_Lacy_Mini_DV	No significant notes.

Workflow summary

Out of all the three workflows, this one is definitely the most troublesome. Whoever is doing the corrections will need to be fairly computer-savvy in order to make corrections in a timely manner. Most of the issues that come with this workflow could be worked through without help of I.T., but one would need to be very familiar with the Windows operating system to do so.

It is helpful to have a set of timecodes that is more-or-less accurate to the cadence of whomever is speaking, as well as a general idea of what someone is saying. The accuracy of the generated transcripts is largely dependant on the quality of the clips. Accuracy of individual transcripts range from about 80% to less than 5%. No matter how good the quality is though, Google doesn't add any form of punctuation to the transcripts, so there will always need to be at least a little correction made.

This workflow serves as a nice middle ground between Rev.com and manual transcription.

Rev.com Correction Notes
(In order of chronological transcription)

Title	Notes
A0095_MaymeWellsLower-MP3	.Rev.com claims a 99.5% accuracy for their services, the document has about 8417 words, so we should expect around 42 errors to be present. Rev.com also doesn't add names to any spoken text, so those will need be identified and added on

	by us. 6 Text omissions 27 Contestable errors 8 Wrong word errors 41 errors total. Right on the money Overall I'm quite impressed, aside from a few small errors, this document looks good Looking at the .vtt file on MovieCaptioner, it seems there is some special character formatting they've implemented for each caption. Other than making the captions a bit harder to read, I don't see them harming us in any way. The .vtt files supplied by Rev.com, when imported into MovieCaptioner and then exported, all the captions appear to be off by a second or two. But this delay isn't reflected in the displayed timecodes in MovieCaptioner. Another issue upon importing causes sequential numbers to appear next to the captions. MovieCaptioner cannot be manipulated in any way to remedy this issue. I was stuck for a while, but found a simple solution to this problem. Instead of using MovieCaptioner to edit the captions, we can use Notepad. If we end up determining Rev.com to be the best workflow, we can finally kill MovieCaptioner once and for all and never use it ever again.
A0938_OLTapp_21_ThoseOtherYears	The first that stands out to me is the fact that this file has cited names for who's speaking. The names are just "Man" and "Woman," but the MaymeWellsLower file hadn't cited any names at all. We could expect about 27 errors to be present. 8 Text Omissions 2 Contestable errors 3 Wrong word errors 13 errors total. Well below the mark. Again, the overall quality is really good. Impressive.
A0212_02_Seeing_is_Believing	This transcription, like all the others, has been of consistently high quality. Out of 4752 words, and

	with a 99.5% accuracy, we can expect there to be about 24 errors. So far, most of the errors are slight word changes, only one error out of the three transcripts I've looked at has been human typing error: "Ar eat" was typed when they mean to type "Are at" 11 Text omissions 0 Contestable errors 6 Wrong word errors 17 errors total. Again, below the expected number.
Workflow summary	
Considering the other two methods, the overall quality and accuracy of the transcripts is fantastic. Rev.com boasts a 99.5% accuracy, and all three clips were either at the mark or well above it. This is the only method of generating transcription where I would be comfortable with no in-house transcriptionist to correct the transcripts. Rev transcript correction only takes a fraction as long as Google transcript correction. This workflow easily beats the other two into the ground.	

For those curious, here are links to the individual transcript reports for each Rev.com file. Each will contain the notes detailed in the table above, as well as a copy of the transcript with highlights pointing out where each of the different error types are.

MovieCaptioner notes

- After working with MovieCaptioner for 5 months, the student is heavily biased against it and this bias should be taken into consideration when viewing all notes about MovieCaptioner.
- For larger files, MovieCaptioner needs a longer amount of time to load a preview of the captions added to the video file. This can take up to three minutes of buffering for clips that are about an hour and a half long.
- MovieCaptioner creates .txt files wherever the .mcp file is located depending on certain properties you apply to the file. For example, MovieCaptioner will always create a .txt file containing important information concerning the formatting of your captions as soon as any captions are added. If you decide to add a translucent background to the captions, a .txt file will be created to provide this. This can cause folders to become very cluttered if you decide to keep all the .mcp files in the same folder, so I'd recommend to keep all the .mcp files in separate folders.
- I don't think MovieCaptioner is the best software to do Google corrections.The user must mouse over to and double click on each and every individual caption they want to edit and then wait for the captions to become editable. For short clips, this wait time is

negotiable. For longer clips the wait time can be almost a second until captions are editable and when there are 1500+ captions to edit, that time adds up quickly.

- MovieCaptioner has some major issues with importing large text files in line form. A good portion of them (maybe one out of ten) do not import the timecode correctly and needs to be fixed manually. There is no discernible pattern and fixing it eats a lot of time, so importing .txt files isn't a great idea.
- When MovieCaptioner imports a WebVTT file, it automatically rounds all of the captions down to the nearest hundredth decimal place. This is strange considering MovieCaptioner has the capacity to export files with timecodes that go down to the thousandth decimal place. It is yet to be determined whether or not this has a significant impact on caption timing.

MovieCaptioner's ability to import files from Rev

- Both .DFXP and .TT files claim to be importable into MovieCaptioner, but cannot be imported. When you select "Import" for either of those options, MovieCaptioner instead only accepts .XML instead of .DFXP or .TT.
- .QT Text files can be imported, but some words don't have spaces between them. The spaces seem to be randomly deleted when imported into MovieCaptioner. This really unusual because MovieCaptioner is built around interacting with Quicktime Media Player, but the made-for-Quicktime caption file doesn't import properly.
- Imported SCC files are near unsalvageable, all the captions are only displayed for less than a second, and you'd have to go through each and every caption to fix it.
- All the .TXT files imported from Rev don't actually add timecodes to the captions, so those are out.
- .SRT files have no issue with importing and exporting

oTranscribe notes

- oTranscribe does not support .AIFF files.
- oTranscribe does support .MP3 files, however, something happens in the process of importing them that causes the timecodes to be very inaccurate. The timecode reflected in oTranscribe is actually 10-15 seconds slower than the true timecode in the original file. .WAV files are completely fine and unaffected by this issue.
- oTranscribe is best utilized for manually creating captions for audio-only files. Video files, google audio correction, and google video correction are best done elsewhere.

Workflow Details

Manual Transcription Workflow (Audio only)

1. Create a new folder with the same name as your audio file. This is where you should keep the raw audio file, as well as exported versions of the transcription.
2. Navigate to https://otranscribe.com
3. Select the “Choose audio (or video) file” button at the top of the page.
4. Navigate to the folder you created containing your audio file.
5. Before you start typing, make sure the transcript window doesn’t have any additional text. (e.g. “Enter your transcript here … Quick tips, etc.)
6. Play the audio and begin transcription. Type down what you hear.
    Here are some helpful keyboard shortcuts
    a. Esc: Play/pause
    b. F1 (or Ctrl+1): Rewind
    c. F2 (or Ctrl+2): Fast-forward
    d. F3 (or Ctrl+3): Slow down
    e. F4 (or Ctrl+4): Speed up
    f. Ctrl+K: Jump to time (brings up window)
    g. Ctrl+0: Return to start
7. At the beginning of each long paragraph or section within the transcript, add a timestamp so it is easy to go to specific points within the recording. Do this by pressing “Ctrl+J” minus the quotes.
8. oTranscribe will save your files in-browser, however this can be unreliable. To export the files to your computer, go to the sidebar to the right of the text editor, and mouse over the bottommost option. The text “Export” should appear next to it. Click on it.
    a. Select “.otr” and “.txt” under the “Download transcript as…” popup window. These two files will serve as backups should oTranscribe not save your progress.
    b. Move the downloaded files to the folder with your raw audio that you created in step 1.
9. To import files, just click on the button right above the “Export” button and navigate to the .otr file you wish to import. oTranscribe will only import .otr files.
10. Once you have finished transcription, export the transcript as “.txt” Move the file to the folder you created in step 1 and rename the file to match the audio file’s name.

Manual Transcription Workflow (Video only)

1. Create a new folder with the same name as your video file. This is where you should keep the raw video file, as well as the MovieCaptioner files.
2. Open a new MovieCaptioner file.
3. Press the “Load Movie” button in the top-right of the MovieCaptioner window.
4. Navigate to the folder containing your video file, select it, and then press “Open”

5. Save the MovieCaptioner file with the same name as your raw video file.
6. Click “Start” in the bottom-left corner to begin adding captions.
7. Audio will loop for 4 seconds. Type what you hear. Hit enter to move onto the next 4 seconds.
    a. You can increase the loop length by one second by pressing “Ctrl+right arrow key”
    b. You can decrease the loop length by one second by pressing “Ctrl+left arrow key”
8. MovieCaptioner will auto save all changes, but it’s best to press “Ctrl+S” to save every once in a while to be sure.
9. Once corrections is finished, export your corrected .vtt file. Go to [Export > WebVTT (File Only)] A new window should pop up.
10. Towards the bottom of the new window, under “File name:” add “.vtt” to the end of the existing file name, minus the quotes.
11. Press “Save”

Google Corrections Workflow

1. Be sure your .vtt file and media you wish to correct are in the same folder on your computer.
2. Open a new MovieCaptioner file.
3. Press the “Load Movie” button in the top-right of the MovieCaptioner window. A window should pop up.
4. Navigate to the folder containing your media file, select it, and then press “Open”
    a. If you are loading an audio file, you will not be able to see it in the navigation window initially. You will need to change the “Files of type:” option in the navigation window. Select “audio/mp3” from the dropdown list and you should be able to select the audio file.
5. Import .vtt file. From the top toolbar, go to [Import > WebVTT]
6. Save the MovieCaptioner file. Make sure you save it in the same folder where your .vtt and media files are.
7. Save the MovieCaptioner file under a new name. (Be sure the new name denotes that this version of the file will be the corrected version)
8. If you have a video file, click the "Preview" button on the bottom of the MovieCaptioner program. This allows you to watch the video as you correct it, as well as ensure the timing is correct on all the captions.
9. If you have an audio file, open the audio file in VLC media player so you can watch the caption timing and correct those as needed.
    a. From the top toolbar, go to [Audio > Visualizations] and then select whichever visualization you prefer. Be sure you have the visualization loaded. You can do this by simply pressing play.
    b. From the top toolbar, go to [Subtitle > Sub Track > Track 1] (If you have multiple .vtt files in the folder that you have your audio file, your desired captions may not

be listed as "Track 1." Keeping your files organized in seperate folders will help mitigate any confusion.

10. As you correct your media, you will need to switch from your media player window and the MovieCaptioner window, editing individual captions as you go. This may be a headache at first, but you will soon be able to switch back and forth with relative ease.
11. Once corrections is finished, export your corrected .vtt file. [Export > WebVTT (File Only)] A new window should pop up.
12. Towards the bottom of the new window, under “File name:” add “.vtt” to the end of the existing file name, minus the quotes.
13. Press “Save”

Rev.com Corrections Workflow (Audio only)

1. Be sure your Rev transcript file and audio file you wish to correct are in the same folder on your computer.
2. Open the Rev transcript file in Notepad.
3. Save the Rev transcript file under a new name. (Be sure the new name denotes that this version of the file will be the corrected version and it is saved in the same folder where your original Rev transcript and audio files are.)
4. Open the audio file in VLC media player so you can watch the caption timing and correct those as needed.
    a. From the top toolbar, go to [Audio > Visualizations] and then select whichever visualization you prefer. Be sure you have the visualization loaded. You can do this by simply pressing play.
    b. From the top toolbar, go to [Subtitle > Sub Track > Track 1] (If you have multiple .vtt files in the folder that you have your audio file, your desired captions may not be listed as "Track 1." Keeping your files organized in seperate folders will help mitigate any confusion.
5. Listening to the audio and watching the subtitles, make any corrections needed to the text or timestamps in Notepad.
6. As you correct your audio, you will need to switch from your media player window and the MovieCaptioner window, editing individual captions as you go. This may be a headache at first, but you will soon be able to switch back and forth with relative ease.
7. Press “Ctrl+S” every once in a while to save your captions in Notepad.
8. Once corrections are finished correcting, go to [File > Save As…] from the top toolbar. A new window should pop up.
9. Towards the bottom of the new window, under “File name:” add “.vtt” to the end of the existing file name, minus the quotes.
10. Right underneath, you’ll find a “Save as type:” option. Click on the down arrow and select “All files” from the dropdown list.
11. Press “Save”

Rev.com Corrections Workflow (Video only)

1. Be sure your Rev transcript file and video file you wish to correct are in the same folder on your computer.
2. Open a new MovieCaptioner file
3. Press the “Load Movie” button in the top-right of the MovieCaptioner window.
4. Navigate to the folder containing your video file, select it, and then press “Open”
5. Import .vtt file. From the top toolbar, go to [Import > WebVTT]
6. Save the MovieCaptioner file. Make sure you save it in the same folder where your .vtt and media files are.
7. Save the MovieCaptioner file under a new name. (Be sure the new name denotes that this version of the file will be the corrected version)
8. Click the "Preview" button on the bottom of the MovieCaptioner program to watch as you correct. (This allows you to watch the video as you correct it, as well as ensure the timing is correct on all the captions.)
9. Begin watching the video and make any corrections as necessary.
10. As you correct your video, you will need to switch from your media player window and the MovieCaptioner window, editing individual captions as you go. This may be a headache at first, but you will soon be able to switch back and forth with relative ease.
11. Once corrections is finished, export your corrected .vtt file. [Export > WebVTT (File Only)] A new window should pop up.
12. Towards the bottom of the new window, under “File name:” add “.vtt” to the end of the existing file name, minus the quotes.
13. Press “Save”